PUERTO RICO AND VIRGIN ISLANDS PRECIPITATION FREQUENCY PROJECT

Update of *Technical Paper No. 42* and *Technical Paper No. 53*

Twentieth Progress Report
1 April 2005 through 30 June 2005

Hydrometeorological Design Studies Center
Hydrology Laboratory

Office of Hydrologic Development
U.S. National Weather Service
National Oceanic and Atmospheric Administration
Silver Spring, Maryland

July 2005

DISCLAIMER

The data and information presented in this report are provided only to demonstrate current progress on the various technical tasks associated with this project. Values presented herein are NOT intended for any other use beyond the scope of this progress report. Anyone using any data or information presented in this report for any purpose other than for what it was intended does so at their own risk

TABLE OF CONTENTS

PUERTO RICO AND VIRGIN ISLANDS PRECIPITATION FREQUENCY PROJECT

Update of *Technical Paper No. 42* and *Technical Paper No. 53*

1. Introduction

The Hydrometeorological Design Studies Center (HDSC), Hydrology Laboratory, Office of Hydrologic Development of NOAA's National Weather Service is updating its precipitation frequency estimates for Puerto Rico and the U.S. Virgin Islands. Current precipitation frequency estimates for the area are contained in *Technical Paper No. 42* "Generalized estimates of probable maximum precipitation and rainfall-frequency data for Puerto Rico and Virgin Islands" (U.S. Weather Bureau, 1961) and *Technical Paper No. 53* "Two- to ten-day rainfall for return periods of 2 to 100 years in Puerto Rico and Virgin Islands" (Miller, 1965). The new project includes collecting data and performing quality control, compiling and formatting datasets for analyses, selecting applicable frequency distributions and fitting techniques, analyzing data, mapping and preparing reports and other documentation.

The project will determine annual precipitation frequencies for durations from 5 minutes to 60 days, for average recurrence intervals from 1 to 1,000 years. The project will review and process all available rainfall data for the Puerto Rico and Virgin Island project area and use accepted statistical methods. The project results will be published as a Volume 3 of NOAA Atlas 14 on the internet (http://www.nws.noaa.gov/ohd/hdsc) with the ability to download digital files.

The project area covers Puerto Rico and the U.S. Virgin Islands of St. Thomas, St. John and St. Croix. The project area is currently divided into 7 regional groups for long duration (24-hour through 60-day) analyses (Figure 1) and 4 regional groups for short duration (60-minute through 12-hour) analyses (Figure 2).

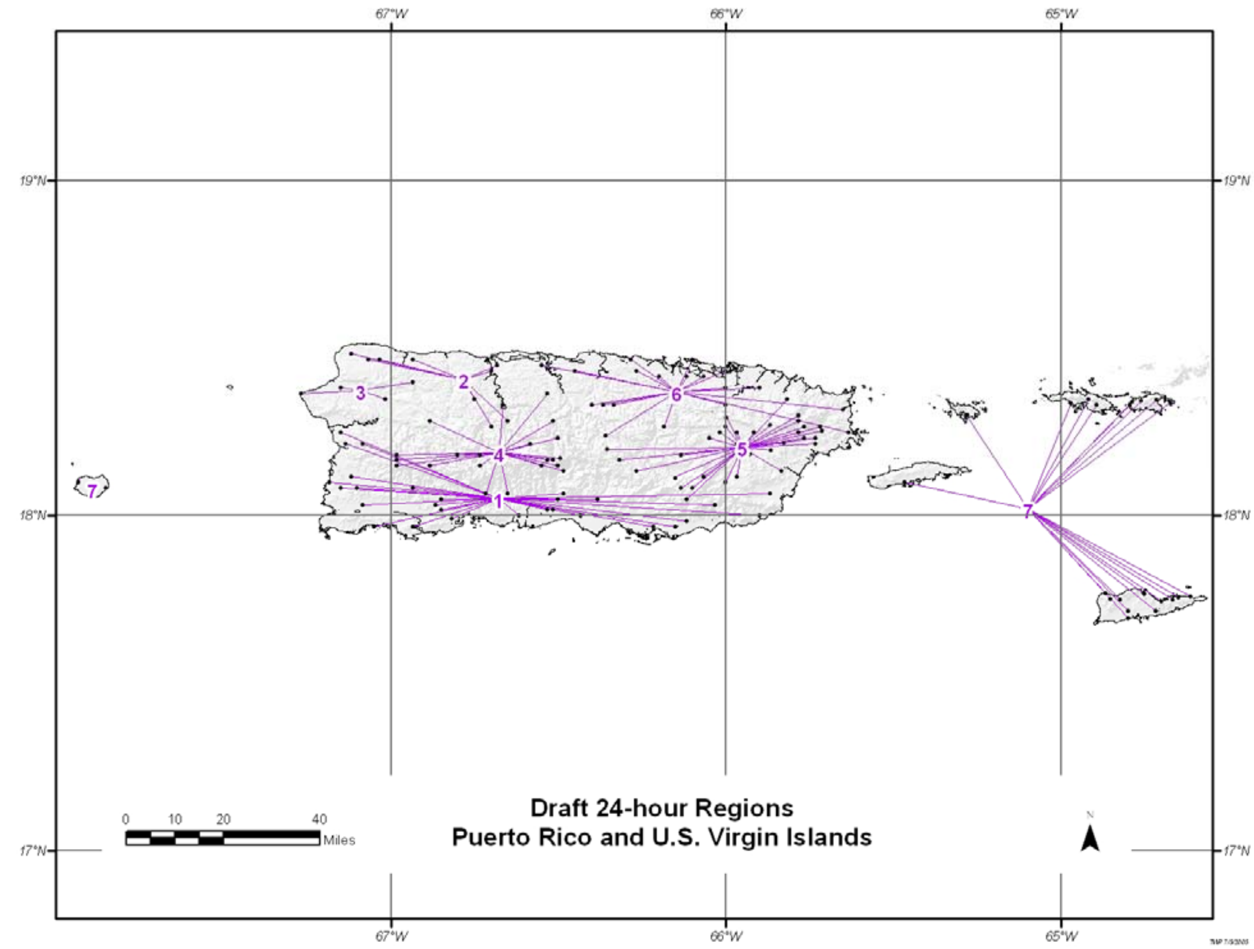


Figure 1. Puerto Rico Precipitation Frequency project area and 7 regional groups based on 24-hour data.

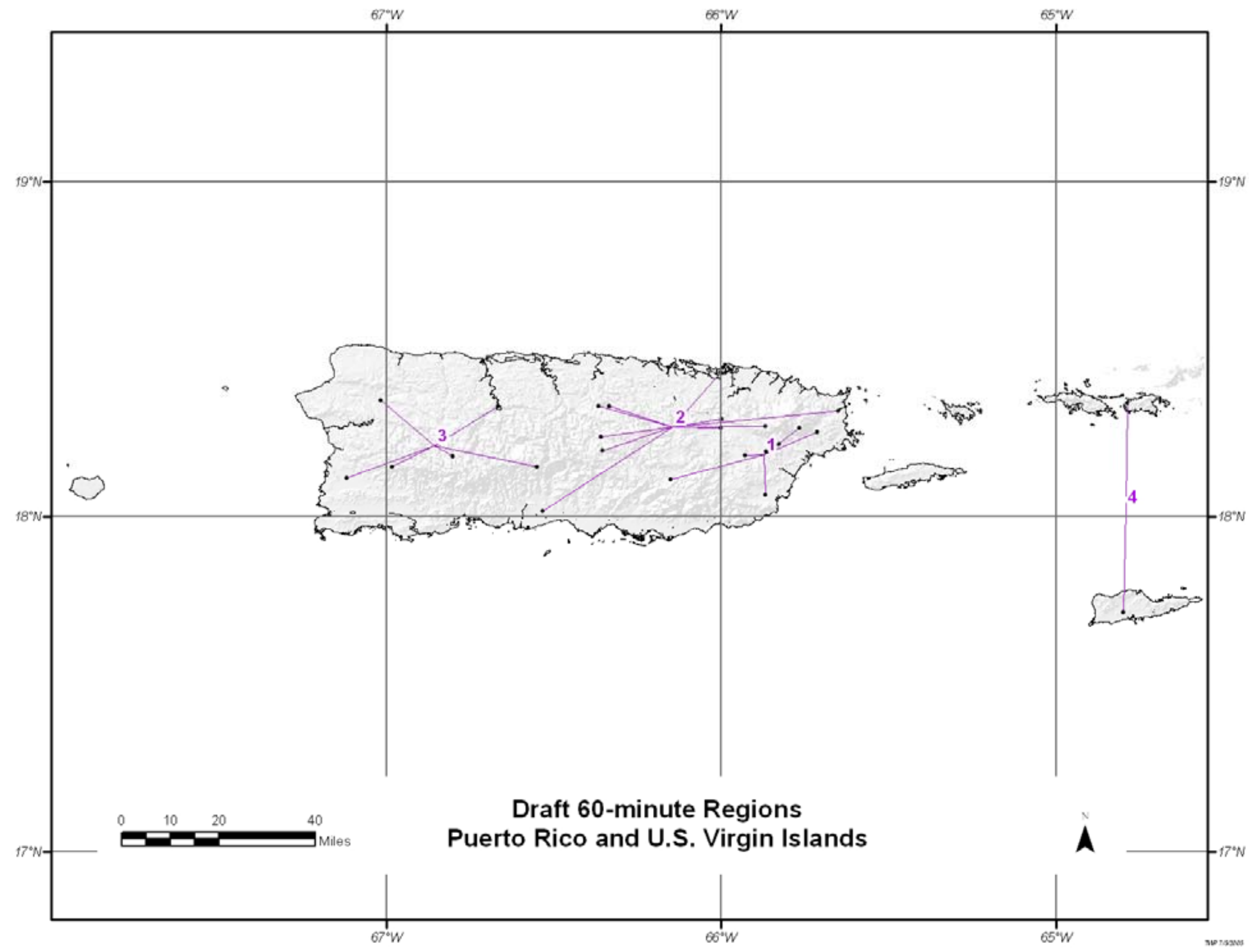


Figure 2. Puerto Rico Precipitation Frequency project area and 4 regional groups based on 60-minute data.

2. Highlights

Collection and quality control of the n-minute dataset was completed. Daily station corrections were made – two stations were deleted, two stations were merged, and suspicious daily data were replaced by 24-hour accumulations where appropriate. Some stations have been identified for additional consideration. *QCseries*, software that screens data series to identify maximum precipitation values that are suspect relative to concurrent data at nearby stations, is being optimized for longer duration data quality control. Additional information is provided in Section 3.1, Data Collection and Quality Control.

The conversion factor for 1-day to 24-hour data was re-assessed and the factor changed slightly from 1.205 to 1.208. There was no change in the 2-day to 48-hour conversion factor. Conversion factors for 1-hour to 60-minute and 2-hour to 120-minute were calculated. The factors are 1.128 and 1.044, respectively. Additional information is provided in Section 3.2, Conversion Factors.

To accurately extract statistically meaningful values, it is ensured that each year has a sufficient number of data, particularly in an assigned “wet season”. The wet seasons for the newest daily regions and hourly regions were defined. Additional information is provided in Section 3.3, Annual Maximum Series Extraction.

Regions based on the 24-hour data were modified to resolve heterogeneity issues and incorporate local climate knowledge (Figure 1). Since there are fewer hourly stations, regions were also delineated based on hourly (i.e., 60-minute) data into four hourly regions (Figure 2). Precipitation frequency estimates were computed using regional L-moments and preliminary maps of the 100-year 24-hour, 1,000-year 24-hour, 100-year 60-minute and 1,000-year 60-minute are being reviewed for spatial anomalies. Regions may be refined based on these results. Since data do not behave ideally in reality, all practical adjustments will be applied to the Puerto Rico precipitation frequency estimates as they were applied in the Ohio River basin project. Finally, an undergraduate student scholar, Precious Lewis, has agreed to work for HDSC on an independent investigation of objective methods to cluster stations into homogeneous regions for precipitation frequency analysis. Additional information is provided in Section 3.4, L-moments and Regionalization.

Based on goodness-of-fit results, climatological considerations and sensitivity testing for all regions in the project area, the Generalized Normal (GNO) distribution was selected to best represent the underlying distributions of the annual maximum data for all 7 daily regions. GNO was also selected for all 4 hourly regions. Additional information is provided in Section 3.5, Frequency Distribution.

On June 9, 2005 HDSC provided the Spatial Climate Analysis Service (SCAS) at Oregon State University with quality-controlled 60-minute and 24-hour mean annual maximum point estimates to be spatially interpolated using PRISM. These means are the basis for the derivation of estimates at all durations and recurrence intervals. Additional information is provided in Section 3.6, Spatial Interpolation.

An analysis of trends in mean and variance and shifts in mean of annual maximum series is nearly complete. Preliminary results are included here. Formal documentation will be included in the final Atlas documentation. Additional information is provided in Section 3.7, Trend Analysis.

The Precipitation Frequency Data Server (PFDS), the on-line portal for all NOAA Atlas 14 deliverables and information, under went several changes. The most significant change was the posting of final documentation and temporal distribution data for Volumes 1 and 2. Additional information is provided in Section 3.8, PFDS.

Progress on the development of areal reduction factors remains slow. Additional difficulties have been encountered in modifying the software to add the analysis of 30-minute and 48-hour durations. Two statistical procedures that will be used to test the differences between the ARF curves generated from the various sites have been applied to the data of two sites. Additional information is provided in Section 3.9, Areal Reduction Factors.

3. Progress in this Reporting Period

3.1 Data Collection and Quality Control

N-minute. N-minute data are precipitation data measured at a temporal resolution of 5-minutes that can be summed to various "n-minute" durations (10-minute, 15-minute, 30-minute, and 60-minute). Automated Surface Observing System (ASOS) data from 1998 through 2004 for San Juan (66-8812) were quality controlled. Extreme 5-minute values above a threshold of 0.5" and accumulated 1-hour values above a threshold of 1" were checked. The availability of additional n-minute stations from ASOS and the National Climatic Data Center (NCDC) was investigated. No additional data was found.

Daily Data. During the work to generate the 1-day to 24-hour conversion factors (see Section 3.2, Conversion Factors), HDSC observed inconsistencies in some concurrent 1-day and 24-hour accumulated maximums. Since it is unreasonable for co-located daily and hourly stations to be inconsistent, the 1-day observation was replaced by the 24-hour accumulated observation that had been converted using the inverse of the conversion factor to a 1-day value for these cases. (See the 19th Progress Report, Section 3.1.3 Daily Data Concurrent 1-day/24-hour Check for more details.)

Several issues related to station location or data were resolved. The data records of the two daily stations 66-0147 and 66-0152 were merged because they were geographically near each other and passed a statistical t-test in which the data were found to be from the same population. Stations 66-6017 (Matrullas Dam) and 66-0849 (Bayney) were deleted due to data quality issues. (See the 19th Progress Report, Section 3.1.3 Daily Data Station Corrections for more details.)

Stations Under Additional Consideration. Some stations have been identified as requiring additional consideration due to their data, period of record and/or consistency with nearby stations. Below is a list of such stations and the situation requiring further scrutiny.

1. 66-8881 – San Sebastian – hourly data

The highest observed 1-hour annual maximum at station 66-8881 (San Sebastian) exceeds the preliminary 1,000-year annual exceedance probability (AEP) estimate. Preliminary investigation indicates the hourly annual maximum data produces a negative L-skewness, a flat upper tail, and therefore little change from the 100-year, 1-hour estimate of 3.28 inches to the 1,000-year estimate of 3.31 inches. The maximum observed 1-hour amount is 3.84 inches. Further investigation into the quality of the 1-hour annual maximum data is therefore required to support this unusual result.

2. 66-9432 – Toro Negro Forest – daily data

The preliminary 1,000-year 24-hour AEP estimate of 60.33 inches at Toro Negro Forest (66-9432) exceeds the Probable Maximum Precipitation (PMP) estimate of about 55 inches from Technical Paper 42. The station has a very high mean annual maximum of 10.05 inches as well as maximum observed 24-hour amount of 28.69 inches (adjusted for clock hour). The relatively short period of record of 22 years at the station leads to very high precipitation frequency estimates. Toro Negro Plant (66-9466) has a longer period of record, 54 years, as well as a lower mean annual maximum of 6.21 inches. Thus, we suspect the short period of record at 66-9432 has resulted in sampling error that has biased the L-moment statistics.

Statistical and sensitivity testing has begun to determine if station 66-9432 can be merged with 66-9466. The two stations are only 1.5 miles apart and about 500 feet different in elevation. 66-9466 has a period of record from 1911 to 1981 and 66-9432 begins in 1982 and continues through the present.

3. Spatial gradient between 66-3431 and 66-1623 – daily data
A surprisingly steep spatial gradient exists between stations Dos Bocas (66-3431) and Caonillas Utuado (66-1623) in a preliminary 100-year 24-hour and 1,000-year 24-hour map generated using Inverse-Distance-Weighting. At the 100-year, 24-hour AEP, the difference between the two stations is 5.70 inches. At the 1,000-year AEP, the difference is more than 15 inches.

The two stations are on a regional boundary that may be, in part, driving the gradient. There are only about 4 miles between these two stations. However, there is 700 feet of elevation difference between the two stations which may account for some of the difference. The period of record at 66-1623 is from 1949 to 1987, while 66-3431 has a much longer period of record from 1937 to the present. Further investigation into the raw precipitation data and metadata of these two stations may provide evidence supporting the gradient.

4. 66-6073 and 66-4702 – daily data
The highest observed annual maximums at Mayaguez City (66-6073) and Isabella Substation (66-4702) exceed the preliminary 1,000-year, 24-hour estimates. Further investigation of the observations at these two stations will be performed before these estimates are accepted.

Longer Durations. Modifications are being made to make the quality control process more efficient by utilizing previously quality controlled data when evaluating longer durations (48-hour through 60-day). The new spatially-based quality control procedure, known as *QCseries*, continued to undergo development, testing and implementation for this purpose. *QCseries* screens data to identify precipitation values that are suspect relative to concurrent data at nearby stations. Code has been added to *QCseries* to cross-reference the events of longer duration time series in the output log file with:

1. a list of previously quality controlled 24-hour extreme events above a certain threshold,
2. the previous duration annual maximum (AMS) or partial duration (PDS) series events, whichever the case may be,
3. the same duration AMS or PDS events.

Flags will indicate when a longer duration contains a 24-hour event that has already been quality controlled or if an event was included in other AMS or PDS data that may have already been quality controlled. Additionally, a new objectively-computed measure of the consistency of the value in question was added in the output log file. This measure is based on the concurrent precipitation at nearby stations and the deviation from spatially distributed values of percent of mean annual precipitation.

QCseries will be run on 4-, 10-, 30- and 60-day durations to evaluate the necessity of the such quality control procedures on longer duration data, since it has typically been assumed that the extensive 24-hour quality control is sufficient to yield good accumulations of longer durations. In addition, a check of statistically discordant stations in the regions, a check for common errors in sequential daily observations, such as erroneously repeating daily values or monthly totals mistakenly entered as a single daily observation on the last day of the month, and a check of low annual maxima will be conducted on longer duration data.

Real-data-check (RDC) flags are occurrences where a maximum observation in the real (observed) data series at a station exceeds a given frequency estimate, in this case the 1,000-year estimate. These stations are being carefully investigated for data quality and appropriate regionalization. Table 1 shows the number of 1,000-year real-data-check flags observed in the preliminary 24-hour and longer duration results for all regions. The number of RDC cases are consistent with statistical expectations and with previous studies. One may expect as many as 6 events that exceed the 1,000-year estimate assuming an average period of record of 50 years and 128 daily stations.

Table 1. Number of observations that exceed the 1,000-year estimates at stations.

Duration	# of 1,000-year RDCs
1-day	2
2-day	0
4-day	0
7-day	1
10-day	2
20-day	2
30-day	3
45-day	3
60-day	4

3.2 Conversion Factors

The conversion factor for 1-day to 24-hour data was re-assessed excluding cases where the 1-day observation was suspect since it was much less than the 24-hour accumulated value during hurricane events. The factor changed slightly from 1.205 to 1.208. There was no change in the 2-day to 48-hour conversion factor.

In order to make hourly and 60-minute data comparable, a conversion is necessary from the constrained ‘clock hour' to unconstrained 60-minute and from 2 hours to 120-

minute. Conversion factors were computed for this project using ratios of the 2-year quantiles computed from annual maxima series and monthly maxima series at one first-order station with co-located hourly and n-minute stations (66-8812 San Juan) and 23 co-located hourly and 15-minute stations with at least 15 years of data (note: 25 15-minute stations with at least 13 years of data were used in the monthly analysis). The time series from concurrent time periods were generated for 60-minute precipitation values summed from n-minute (or 15-minute) observations and for co-located hourly observations. The series were analyzed separately using L-moments and four distributions (GEV, GLO, GNO, and GPA). Ratios of 2-year 60-minute to 2-year 1-hour quantiles were generated and averaged. The conversion factor was further verified by a regression analysis of concurrent annual maxima pairs and monthly maxima pairs. There were 694 concurrent annual maxima data pairs from 24 co-located stations and 7,836 concurrent monthly maxima pairs from 26 co-located stations. The resulting conversion factor was 1.128 for 1-hour to 60-minute and 1.044 for 2-hour to 120-minute. This is in close agreement with Technical Paper 42 which used 1.13 for the 1-hour to 60-minute conversion and with NOAA Atlas 14 Volumes 1 and 2 (see Table 2). No conversion was provided for 2-hour to 120-minutes in Technical Paper 42, but this factor is in close agreement with NOAA Atlas 14 Volumes 1 and 2.

Table 2. Conversion factors for constrained to unconstrained observations.

	Conversion Factors			
Project	**1-day to 24-hour**	**2-day to 48-hour**	**1-hour to 60-minute**	**2-hour to 60-minute**
NOAA Atlas 14 Vol. 1 (Semiarid Southwestern United States)	1.14	1.03	1.12	1.03
NOAA Atlas 14 Vol. 2 (Ohio River Basin and Surrounding States)	1.13	1.04	1.16	1.05
NOAA Atlas 14 Vol. 3 (Puerto Rico and U.S. Virgin Islands)	**1.208**	**1.134**	**1.128**	**1.044**

3.3 Annual Maximum Series Extraction

Wet Season Criteria. The extraction software requires that at least 50% of the months in the wet season be present to allow an annual maximum to be extracted for a given year. The wet season for each location is assigned by assessing histograms of annual maximum precipitation for each homogeneous region (see the 19th Progress Report, Section 3.2, Annual Maximum Series Extraction for more details). In this reporting period, regions for the 24-hour and 60-minute analyses were refined (see Section 3.4, L-moments and Regionalization), so it was necessary to redefine the wet seasons. The wet seasons for the newest daily regions and hourly regions are shown in Tables 3 and 4.

Table 3. Wet season months based on 1-day data.

Region	Start month	End month
1	5	11
2	4	12
3	4	11
4	4	11
5	5	11
6	4	12
7	4	11

Table 4. Wet season months based on 1-hour data.

Region	Start month	End month
1	4	11
2	4	11
3	4	11
4	1	12

3.4 L-moments and Regionalization

Regionalization Results. Regions based on the 24-hour data were revised to resolve heterogeneity issues and incorporate local climate knowledge. In April, a conference call with Israel Matos and Ernesto Morales was held to discuss local climate and regionalization. The resulting seven regions are shown in Figure 1 and the number of stations and H1 measure are shown in Table 5. The heterogeneity measure, H1, tests between-station variations in sample L-moments for a group of stations with what would be expected for a homogeneous region based on coefficient of L-variation (Hosking and Wallis, 1997). An H1 measure greater than 2 (H1>2) indicates heterogeneity and H1<2 indicates homogeneity. All regions are now homogeneous.

Table 5. Heterogeneity results for 24-hour regions in Figure 1.

region	# daily stations	# hourly stations	# total stations	H1
1	27	3	30	1.62
2	9	0	9	1.66
3	4	1	5	1.51
4	21	4	25	1.26
5	26	10	36	1.19
6	18	5	23	0.64
7	23	2	25	0.17
total	**128**	**25**	**153**	

Since there are fewer hourly stations, regions were also delineated based on hourly (i.e., 60-minute) data. There are four hourly homogeneous regions (Figure 2). The number of stations and H1 measure are shown in Table 6.

Table 6. Heterogeneity results for 60-minute regions in Figure 2.

region	**# hourly stations**	**H1**
1	7	-1.30
2	10	1.60
3	6	0.37
4	2	-0.05
total	**25**	

Precipitation frequency estimates were computed using regional L-moments and preliminary maps of the 100-year 24-hour, 1,000-year 24-hour, 100-year 60-minute and 1,000-year 60-minute are being reviewed for spatial anomalies such as steep gradients and/or bull's eyes. Regions may be refined based on these results.

In addition, all statistically discordant stations and 1,000-year real-data-check (RDC) cases within a region were checked for data quality. [Again, real-data-check (RDC) flags are occurrences where a maximum observation in the real (observed) data series at a station exceeds a given frequency estimate, in this case the 1,000-year estimate.] There were no 1,000-year RDCs in the 60-minute results. There were two cases in the 24-hour data (21.30" on September 22, 1998 at station 66-6073 Mayaguez City and 11.20" on August 9, 1899 at station 66-4702 Isabela Substation).

Practical Adjustments. Once the regions are determined, all practical adjustments will be applied to the Puerto Rico precipitation frequency estimates during the L-moment analysis as they were applied in the Ohio River basin project. Practical adjustments must be applied because in reality, data do not always behave ideally. Nor are datasets always collected perfectly through time or in dense spatial networks. Since quantiles for each duration and station in this project are computed independently, practical adjustments are applied to produce realistic final results that are consistent in duration, frequency and space. The applied adjustments are briefly described below:

1. An annual maxima adjustment ensures consistency from one duration to the next longer duration for each given year at all daily stations.
2. Since hourly and daily durations are computed separately and from different data sets, it is necessary to ensure consistency of precipitation frequency estimates through the durations at co-located daily and hourly stations.
3. To ensure that hourly-only stations are consistent with nearby co-located hourly/daily stations that occur in different regions and to reduce spatial bull's eyes that may be observed in hourly results, an adjustment is applied to hourly-only stations.
4. Since the quantiles of each duration at a given station are calculated separately, inconsistencies could occur where a shorter duration had a quantile that is higher than the next longer duration at a given average recurrence interval. Internal consistency adjustments are applied to mitigate this.

Please see pages 38-40 of NOAA Atlas 14 Volume 2.1 for additional details (http://hdsc.nws.noaa.gov/hdsc/pfds/docs/NA14Vol2_4analysis.pdf).

Objective Regionalization. An undergraduate student scholar, Precious Lewis, has agreed to work with HDSC through the NOAA Educational Partnership Program from June 13, through August 5, 2005. She will conduct an independent investigation of objective methods to cluster stations into homogeneous regions for precipitation frequency analysis possibly using IDRISI Kilimanjaro software from Clark Labs. IDRISI software is a GIS and Image Processing software solution that includes functions for the analysis of digital spatial information. She will use Puerto Rico data to test the method. Her work will provide insight into the current regionalization process and provide a more efficient means to arrive at final regions. If feasible, the developed process will provide stronger initial regions for future projects. These initial regions would still be carefully scrutinized and modified where necessary so that they meet the needs of the analysis.

3.5 Frequency Distribution

It is assumed that the stations within a region share the same shape but not scale of their precipitation frequency distribution curves. It is not assumed that these factors or the distribution itself are common from region to region. In other words, a probability distribution is selected and its parameters are calculated for each region separately.

Since a three-parameter distribution, which behaves both relatively reliably and flexibly, is more often selected to represent the underlying population, candidate theoretical distributions included: Generalized Logistic (GLO), Generalized Extreme Value (GEV), Generalized Normal (GNO), Generalized Pareto (GPA), and Pearson Type III (PE3). The five-parameter Wakeby distribution would have been considered only if the three-parameter distributions were found unsuitable for a region, but this did not happen. Three goodness-of-fit measures were computed to assist in selecting the most appropriate distribution for the region. These were the Monte Carlo Simulation test, real-data-check test, and RMSE of the sample L-moments (Please see Section 4.5 Choice of Frequency Distribution in NOAA Atlas 14 documentation of Methods-Analysis at http://hdsc.nws.noaa.gov/hdsc/pfds/docs/NA14Vol2_4analysis.pdf for additional details about the three tests.)

A final decision of the most appropriate distribution for a region was primarily based upon a summary of the three tests. The goodness-of-fit tests were done on a region-by-region basis. The results from the three tests provide a strong statistical basis for selecting the most appropriate distribution. However, the goodness-of-fit results were then weighed against climatologic and geographic consistency considerations. During a review of results on a macro-scale, the distribution identified by the three tests may have been changed to reduce bull's eyes and/or gradients in precipitation frequency estimates between regions. An effort was also made to maintain consistency of selected distribution from region to region. The use of an alternate distribution was supported with sensitivity testing to ensure that results using the selected distribution are acceptable (i.e., changes in 100-year quantiles were less than 5%).

Tables 7 and 8 show the preliminary results of these tests and the selected distribution for the 24-hour data and 60-minute data, respectively. Based on the goodness-of-fit results, climatological considerations and sensitivity testing for all regions in the project area, GNO was selected to best represent the underlying distributions of the annual maximum data for all 7 daily regions. GNO was also selected for all 4 hourly regions.

The best-fitting frequency distribution results are similar to those found in NOAA Atlas 14 Volumes 1 and 2, thus providing support to the distribution selection method for this project area even though it experiences a different climate that is highly influenced by tropical rainfall.

Table 7. Preliminary goodness-of-fit test results for 24-hour annual maximum series data in each daily region.

		Monte Carlo Simulation		Real-data-check test		RMSE test		
region	rank	Distribution	test value	Distribution	test value	distribution	RMSE	selected
1	1st	GNO	-0.68	GNO	18.0	GNO	0.11065	GNO
	2nd	GEV	1.22	GEV	17.0	GEV	0.11327	
	3rd	GLO	3.15	PE3	14.5	GPA	0.11563	
2	1st	PE3	0.22	PE3	17.0	PE3	0.15636	GNO
	2nd	GNO	1.48	GNO	16.0	GNO	0.15758	
	3rd	GEV	2.09	GEV	16.0	GEV	0.15884	
3	1st	GLO	0.56	GEV	20.0	GNO	0.25259	GNO
	2nd	GEV	-0.61	GNO	17.5	GEV	0.25299	
	3rd	GNO	-1.10	PE3	15.0	GLO	0.25396	
4	1st	GLO	0.28	GNO	19.0	GNO	0.14004	GNO
	2nd	GEV	-0.53	GEV	17.0	GEV	0.14005	
	3rd	GNO	-2.05	GPA	14.5	GLO	0.01429	
5	1st	GNO	0.32	GNO	20.0	GNO	0.11850	GNO
	2nd	GEV	2.09	PE3	17.0	GEV	0.12234	
	3rd	PE3	-2.71	GEV	15.5	GPA	0.12271	
6	1st	PE3	0.62	PE3	21.0	GPA	0.12216	GNO
	2nd	GNO	2.86	GPA	21.0	PE3	0.12341	
	3rd	GPA	-3.03	GNO	16.0	GNO	0.12721	
7	1st	GNO	-0.23	PE3	18.5	GNO	0.12378	GNO
	2nd	GEV	0.96	GNO	18.0	GEV	0.12754	
	3rd	GLO	2.20	GPA	15.5	GPA	0.12789	

Table 8. Preliminary goodness-of-fit test results for 60-minute annual maximum series data in each hourly region.

region	rank	Monte Carlo Simulation		Real-data-check test		RMSE test		selected
		Distribution	test value	Distribution	test value	distribution	RMSE	
1	1st	GEV	-0.03	PE3	19.0	GEV	0.21216	GNO
	2nd	GNO	-0.50	GNO	18.0	GNO	0.21483	
	3rd	GLO	0.93	GLO	14.0	GPA	0.21848	
2	1st	PE3	1.39	PE3	18.0	PE3	0.14649	GNO
	2nd	GEV	1.54	GNO	18.0	GEV	0.14660	
	3rd	GNO	1.71	GEV	18.0	GNO	0.14748	
3	1st	GNO	-0.37	PE3	18.5	GEV	0.19140	GNO
	2nd	PE3	-0.49	GNO	18.5	GNO	0.19146	
	3rd	GEV	-0.62	GEV	18.5	PE3	0.19214	
4	1st	GEV	0.00	PE3	16.0	GNO	0.23590	GNO
	2nd	PE3	-0.02	GNO	16.0	PE3	0.23594	
	3rd	GNO	0.06	GLO	16.0	GEV	0.26305	

3.6 Spatial Interpolation

On June 9, 2005 HDSC provided the Spatial Climate Analysis Service (SCAS) at Oregon State University with quality-controlled 60-minute and 24-hour mean annual maximum (a.k.a. "Index flood") point estimates. There are 25 60-minute stations and 137 24-hour stations (which includes daily and hourly stations). SCAS will use PRISM (Parameter-Regression Independent Slopes Model) (Daly et al. 1994, 2002) to spatially distribute these means to high-resolution (90-meter) grids, which will later be used by HDSC as the basis for deriving the precipitation frequency grids.

3.7 Trend Analysis

An analysis of trends in mean and variance and shifts in mean of the annual maximum series is nearly complete. Preliminary results are shown here. Formal documentation will be included in the final Atlas documentation.

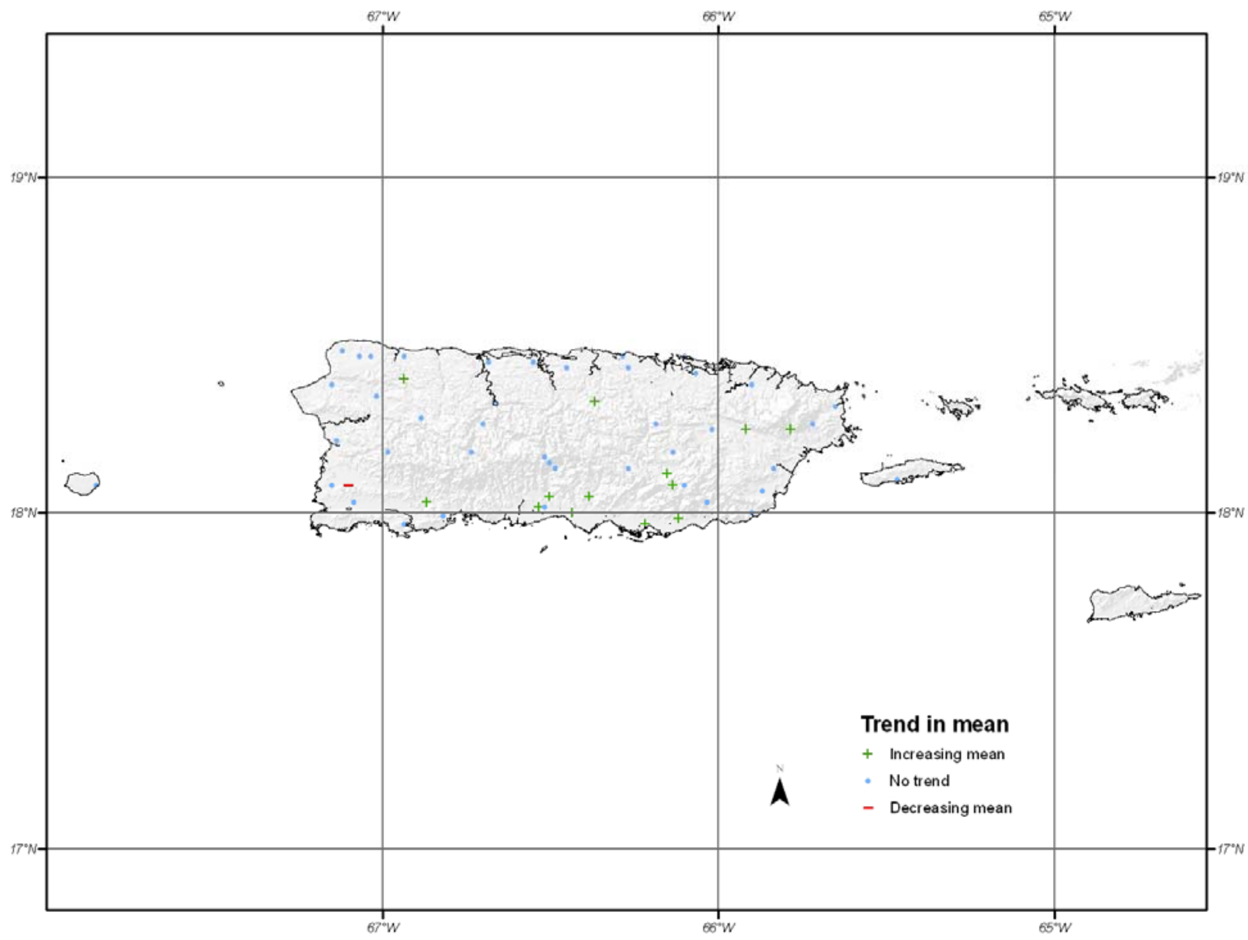
Linear Trend Results. Linear trend tests were conducted to determine if there were any general increasing or decreasing patterns in the 1-day annual maximum series at a station through time. Data were tested for a linear trend in annual maximum series using the linear regression model and t-test of the correlation coefficient (Maidment, 1993, p17.30) at the 90% confidence level. Linear trends in variance were also tested by constructing a variance-related variable, an index of the square of deviation, or $v_i = (x_i - \bar{x})^2$ where, x_i is the annual maximum series data for i = 1, 2, …, n - the data year at a station, and $\bar{x}$ is the mean of the data. The index was then applied as a simple variable in the linear trend model. It was necessary for there to be a continuous time series to be eligible for the linear trend test. A minimum length of 50 years was chosen because it was sufficient to give reliable results and was close to the average data length of available stations.

Of 128 stations in the project area, 55 (or 43.0%) were eligible for the test. Of those tested stations, 25.5% exhibited a linear trend in their annual maximum series (23.6% in a positive direction, 1.8% in a negative direction). It is notable that of the stations exhibiting a linear trend, only 1 station was in the negative direction. This station, San German 4 W, had only 67 years of data and the trend seemed to be driven by 3 high events before 1930. Table 9 lists the linear trend results in the project area. Figure 3 shows the spatial distribution of stations with linear trends.

Table 9. Number of stations tested and linear trend test results in the Puerto Rico project area.

	# Tested	# No Trend	# Trend	# Pos. Trend	# Neg. Trend	% tested with Trend
Total	**55**	**41**	**14**	**13**	**1**	**25.5**

Figure 3. Spatial distribution of linear trend results, where "+" indicates a station with a positive trend and "-" indicates a negative trend.

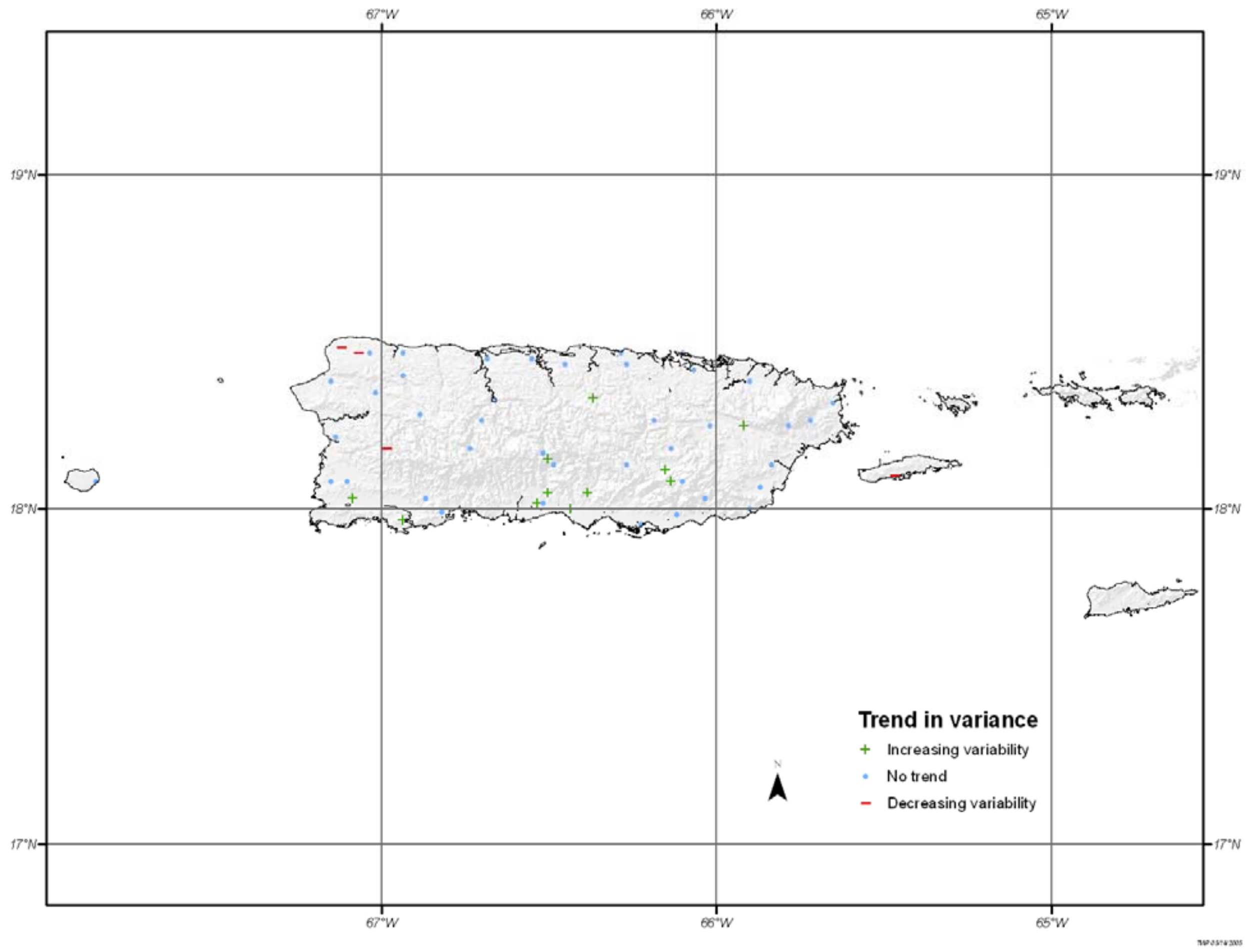


Of the 55 stations tested, 27.3% exhibited a trend in the variance of annual maximums (20.0% in a positive direction, 7.3% in a negative direction). In other words, 20.0% of the stations that exhibited such a trend showed an increase in variance. Table 10 lists the trend in variance results in the project area. Figure 4 shows the spatial distribution of those stations that had a trend in variance.

Table 10. Number of stations tested and linear trend in variance test results by state.

	# Tested	# No Trend	# Trend	# Pos. Trend	# Neg. Trend	% tested with Trend
Total	**55**	**40**	**15**	**11**	**4**	**27.3**

Figure 4. Spatial distribution of trend in variance results, where "+" indicates a station with a positive trend and "-" indicates a negative trend.



HDSC also briefly investigated whether the occurrence and characteristics of hurricanes at stations in Puerto Rico influenced trends in annual maximum precipitation. The findings were inconclusive. Some stations showed that hurricane events (along with the extreme January 1992 event) drove the trends while other stations did not.

Shift Test Results. A shift test was conducted to compare the means of 1-day annual maximum series for two consecutive time periods at a station. The data were tested for shifts in mean using Mann Whitney non-parametric test (Newbold, 1988, p403) and the t-test (Lin, 1980, p160) at the 90% confidence levels. The Mann Whitney is a qualitative test that indicated if a shift occurred but not the direction of the shift. The t-test provided a quantitative measurement of the percentage that the mean shifted from one time period to the next. Both tests give consistent results suggesting that the parametric t-test results can be used with assurance to assign quantitative values to observed shifts. A division of 1958 was tested because 1958 was the final year for which Technical Paper 42 (U.S. Weather Bureau, 1961) had data. The results would indicate whether a shift has occurred since the publication of earlier precipitation frequency estimates. A minimum of 30 years of data in each data segment were required at a station to test for shifts in mean.

Since the Mann Whitney test uses ranks, it was better to have similar sizes between the two subsamples. A threshold of 30 years difference in the lengths of the subsamples was set based on testing and used to screen the stations eligible for that test. However, since the t-test is a parametric test following the t-distribution or Normal distribution, the test is less sensitive to the difference between the sample sizes. In this project, there were 20 stations that were screened out (not eligible) for the Mann Whitney test that were included for the t-test.

The results when using 1958 as the division were:

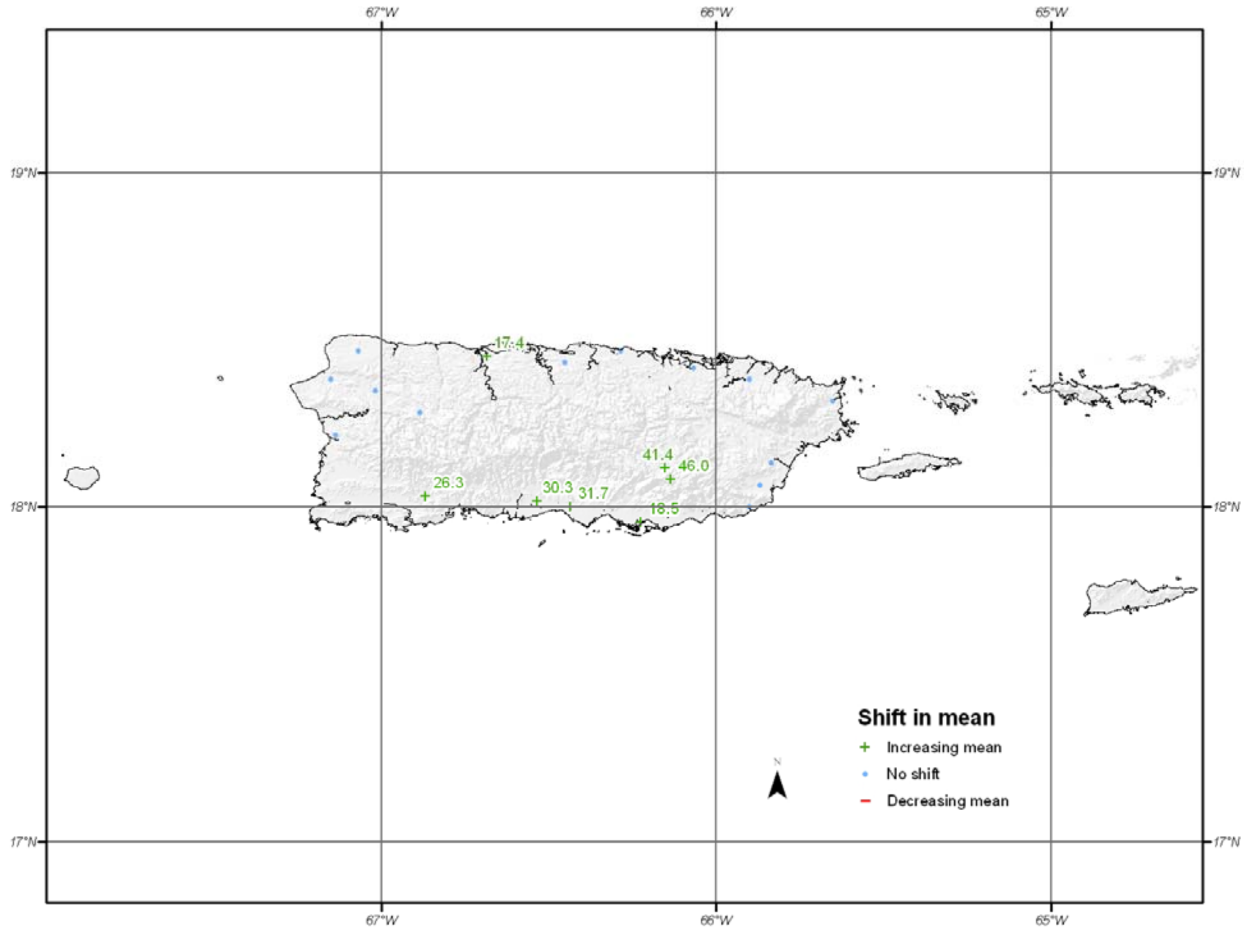
- T-test: 22 of 128 (17.2%) were eligible. 36.4% of those tested had a shift in mean (36.4% increased in mean, 0.0% decreased in mean).
- Mann Whitney test: 22 of 128 (17.2%) were eligible. 36.4% of those tested had a shift in mean.

Table 11 lists the shift in mean results comparing pre-1958 data and post-1958 data in the project area. The last column in the table shows the average percent change in mean for the project area. Overall, the majority of stations tested did not exhibit a shift in mean. Where shifts did occur, the shifts in mean showed a clear preference toward increasing shifts. Figure 5 shows the spatial distribution of the stations that have a shift in mean. The numbers plotted above the station location indicate the percentage of change in mean at each station.

Table 11. Number of stations tested and test for shift in mean results (1958 split) by state.

	# Tested	# No Shift	# Shift	# Pos. Shift	# Neg. Shift	% Change in Mean
Total	**22**	**14**	**8**	**8**	**0**	**36.4 (avg)**

Figure 5. Spatial distribution of shift in mean results, where "+" indicates a station with a positive trend, "-" indicates a negative trend and the number indicates the percentage of change (1958 split).



3.8 PFDS

The Precipitation Frequency Data Server (PFDS) - the on-line portal for all NOAA Atlas 14 deliverables and information – under went a few changes. The most significant change was on the “NOAA Atlas 14 Temporal Distributions” page, where users can now download the data (as comma-delimited files) used to plot the temporal distribution graphs in the NOAA Atlas 14 documentation (Appendix A.1). Additionally, the “NOAA Atlas 14 Version Tables” and “PFDS and NOAA Atlas 14 Documentation” pages were modified to include the release of edited final documentation for Volume 1 and final documentation for Volume 2.

HDSC continuously monitors the hits, integrity and performance of the PFDS, which continues to receive an increasing number of hits per month. The graph (Figure 6) below summarizes the number of individual data inquires made since January 2004, while the map (Figure 7) indicates the locations of inquires during the past quarter.

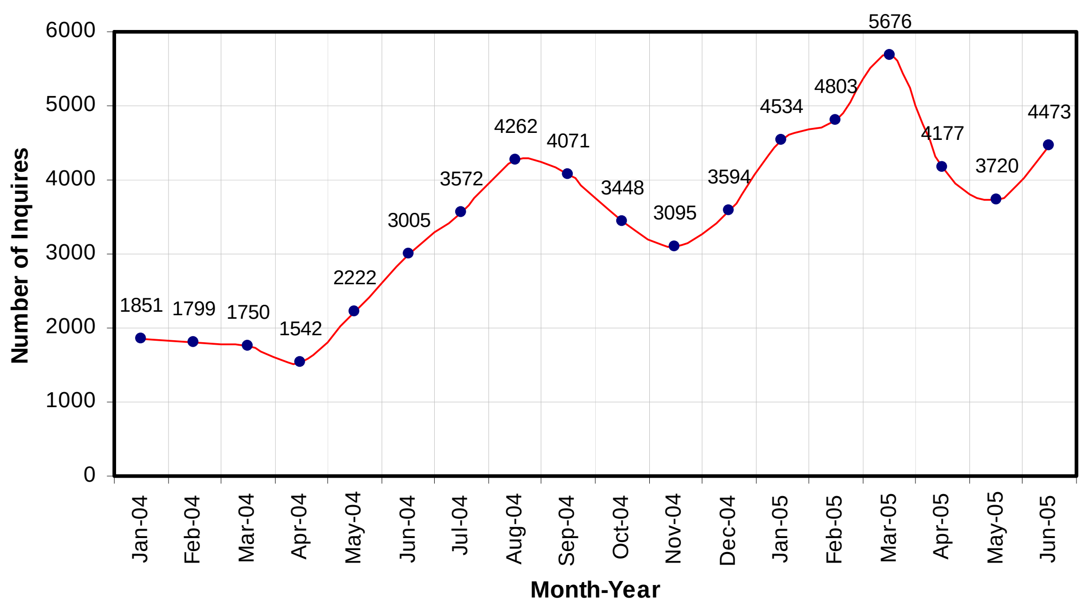


Figure 6. Number of individual PFDS data inquires per month.

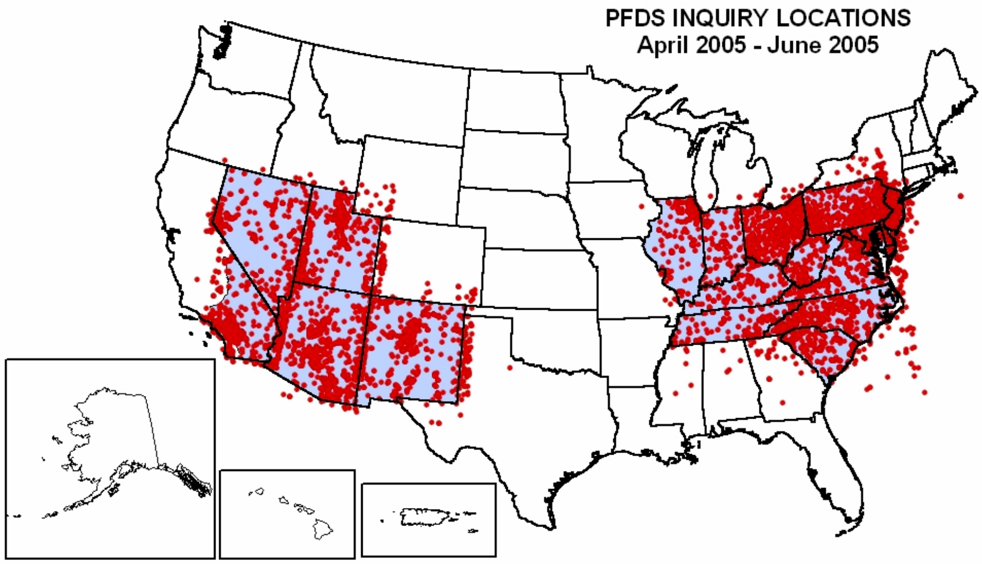


Figure 7. Map of 12,370 PFDS data inquiry locations during the period April-June 2005.

3.9 Areal Reduction Factors

Work continues in the development of geographically-fixed Areal Reduction Factor (ARF) curves for basin area sizes of 10 to 400 square miles. Progress has been slow due to difficulties in completing the software related to the addition of two durations (30-minutes and 48-hours). The use of Hydrometeorological Automated Data System (HADS) data is being investigated in order to supplement other precipitation data with the goal of possibly adding two basin sites (in the New York-New Jersey area and in Oklahoma) to the ARF study.

Two statistical and objective testing procedures, the sign test (Himmelblau, 1970) and a modified "student t" test (Siegel, 1961) will be applied to detect differences among the ARF curves at the various sites. Results of these tests on the data from two sites are being examined.

4. Issues

4.1 Recent Award

On May 13th, 2005, Geoff Bonnin, Director of HDSC, was awarded the 2005 Administrator's Award from Conrad C. Lautenbacher, Jr., Vice Admiral, U.S. Navy (Ret.), Under Secretary of Commerce for Oceans and Atmosphere for "exceptional leadership in developing climatic precipitation information for engineering design standards for the nation's construction and insurance industry." The Administrator's Award is an extremely high level of recognition within NOAA. Mr. Bonnin's honor reflects the accomplishments of entire staff during the past few years.

4.2 1-year Precipitation Frequency

HDSC has been approached by the State of Maryland State Highway Administration (MDSHA) to calculate and include the 1-year average recurrence interval (ARI) precipitation frequency estimates for NOAA Atlas 14 Volume 2. Discussions are being held with MDSHA on funding and contractual mechanisms. We anticipate that the additional estimates will be computed for the entire Volume 2 domain as it is cheaper to perform this calculation than to segregate out a calculation for Maryland alone.

5. Projected Schedule and Remaining Tasks

The following list provides a tentative schedule with completion dates. Brief descriptions of tasks to be worked on are also included in this section.

Data Collection and Quality Control [July 2005]
Trend Analysis [July 2005]
Temporal Distributions of Extreme Rainfall [July 2005]
L-Moment Analysis/Frequency Distribution [July 2005]
Peer Review of Spatially Interpolated Point Estimates [August 2005]
Spatial Interpolation of Grids [September 2005]
Precipitation Frequency Maps [November 2005]
Web Publication [October 2005]
Spatial Relations (Areal Reduction Factors) [September 2005]

5.1 Data Collection and Quality Control

During the next quarter, quality control of the longer (2-day through 60-day) and shorter (2-hour through 12-hour) will be completed. N-minute ratios will be calculated.

5.2 L-Moment Analysis/Frequency Distribution

A comprehensive L-moment statistical analysis will be completed for all durations and regionalization will be finalized.

5.3 Trend Analysis and Temporal Distributions

The analysis for trends in the annual maximum time series will be finalized and an analysis of the hourly temporal distributions of heavy rainfall will be completed.

5.4 Cross-correlation

1-day annual maximum series data will be investigated for cross correlation between stations to assess intersite dependence, since it is assumed for precipitation frequency analysis that events are independent.

5.5 Spatial Interpolation

Pending receipt of the mean annual maximum grids from SCAS, HDSC will generate the grids for 24-hour and 60-minute grids. A Peer Review of the 60-minute and 24-hour mean annual maximum grids and 100-year precipitation frequency grids will then commence. After addressing reviewer concerns, HDSC will begin generating the grids for all durations and recurrence intervals.

5.4 Areal Reduction Factors (ARF)

Computations for the ARF curves will be completed in the next quarter for 14 areas. The resulting curves will be tested for differences to determine if a single set of ARF curves is applicable to the entire U.S. or whether curves vary by region.

References

Arkell, R.E., and F. Richards, 1986: Short duration rainfall relations for the western United States, *Conference on Climate and Water Management-A Critical Era and Conference on the Human Consequences of 1985's Climate*, August 4-7, 1986. Asheville, NC.

Bell, F.C., 1976: The areal reduction factors in rainfall frequency estimation, *Natural Environmental Research Council (NERC), Report No. 35*, Institute of Hydrology, Wallingford, U.K. 25pp.

Bonnin, G., D. Todd, T. Parzybok, B. Lin, D. Riley, and M. Yekta, 2004: Precipitation frequency atlas of the United States. *NOAA Atlas 14 Volume 1*, Silver Spring, Maryland. http://hdsc.nws.noaa.gov/hdsc/.

Carter, M. M. and J. B. Elsner, 1996: Convective Rainfall Regions of Puerto Rico, *International Journal of Climatology*, **16**, pp. 1033-1043.

Carter, M. M., S. Bennett, and J.B. Elsner, 1997: Monthly Rainfall Climatology for Puerto Rico, *NWS Southern Topics February 1997*.

Colon, J, 1966: On the Mechanisms for Production of Rainfall in Puerto Rico, Weather Bureau *Technical Memorandum 15*, San Juan, Puerto Rico.

Daly, C., W.P. Gibson, G.H. Taylor, G.L. Johnson, and P.Pasteris, 2002: A knowledge-based approach to the statistical mapping of climate. *Climate Research*, **23**, 99-113.

Daly, C., R.P. Neilson, and D.L. Phillips, 1994: A Statistical-Topographic Model for Mapping Climatological Precipitation over Mountainous Terrain. *Journal Applied. Meteorology*, **33**, 140-158.

Frederick, R.H. and J.F. Miller, 1979: Short Duration Rainfall Frequency Relations for California, Third Conference on Hydrometeorology, August 20-24, 1979. Bogata Columbia.

Frederick, R.H., V.A. Myers and E.P. Auciello, 1977: Five- to 60-minute precipitation frequency for the eastern and central United States, *NOAA Technical Memo. NWS HYDRO-35*, Silver Spring, MD, 36 pp.

Hershfield, D.M., 1961: Rainfall frequency atlas of the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years, *Weather Bureau Technical Paper No. 40*, U.S. Weather Bureau. Washington, D.C., 115 pp.

Himmelblau, D.M., 1970: “Sign Test for Median Difference in Paired Observations”, *Process Analysis by Statistical Methods*, page 68.

Hosking, J.R.M. and J.R. Wallis, 1997: *Regional frequency analysis, an approach based on L-moments*, Cambridge University Press, 224 pp.

Huff, F. A., 1990: Time Distributions of Heavy Rainstorms in Illinois, *Illinois State Water Survey*, Champaign, 173, 17pp.

Institution of Engineers, Australia, 1987: *Australian Rainfall and Runoff, 3rd Edition*, The Institution of Engineers, Australia. Canberra.

Lin, B. and L.T. Julian, 2001: Trend and shift statistics on annual maximum precipitation in the Ohio River Basin over the last century. Symposium on Precipitation Extremes: Prediction, Impacts, and Responses, 81st AMS annual meeting. Albuquerque, New Mexico.

Lin, Shao-Gong, 1980: *Basic Probability and Statistics*. People's Education Publisher, Beijing, China, 162 pp.

Maidment, D. R., 1993: *Handbook of Hydrology*. McGraw-Hill Publishing, 29.47 pp.

Malgren, B., and A. Winter, 1999: Climate Zonation in PR based on Principal Component Analysis and an Artificial Neural Network, *Journal of Climate*, **12**, pp. 977-985.

Miller, J.F., 1964: Two- to ten-day precipitation for return periods of 2 to 100 years in the contiguous United States, *Technical Paper No. 49*, U.S. Weather Bureau and U.S. Department of Agriculture, 29 pp.

Miller, J.F., 1965: Two- to Ten-Day Rainfall for Return Periods of 2 to 100 Years in Puerto Rico and the Virgin Islands, *Technical Paper No. 53*, U.S. Weather Bureau and U.S. Department of Agriculture, 35 pp.

Miller, J.F., R.H. Frederick and R.J. Tracy, 1973: Precipitation-frequency atlas of the western United States, *NOAA Atlas 2*, 11 vols., National Weather Service, Silver Spring, MD.

Myers, V.A. and R.M. Zehr, 1980: A Methodology for Point-to-Area Rainfall Frequency Ratios, *NOAA Technical Report NWS 24*, Office of Hydrology, National Weather Service, Silver Spring, MD.

Newbold, P., 1988: *Statistics for Business and Economics*. Prentice Hall, 866 pp.

Spiegel, M.R., 1961: "Tests of Significance Involving Sample Differences", *Theory and Problems of Statistics*, pages 170-171.

U.S. Weather Bureau, 1961: Generalized Estimates of Probable Maximum Precipitation and Rainfall-Frequency Data for Puerto Rico and Virgin Islands for Areas to 400

Square Miles, Durations to 24 Hours and Return Periods from 1 to 100 Years, *Technical Paper 42*, Washington, DC, 94pp.